


Article

Investigating the Diffusion of Workload-Induced Stress—A Simulation Approach

Niyi Ogunbiyi * , Artie Basukoski and Thierry Chausalet

School of Computer Science and Engineering, University of Westminster, London W1W 6UW, UK;
A.Basukoski@westminster.ac.uk (A.B.); chausst@westminster.ac.uk (T.C.)

* Correspondence: oluniyi.ogunbiyi@my.westminster.ac.uk

Abstract: Work-induced stress is widely acknowledged as harming physical and psychosocial health and has been linked with adverse outcomes such as a decrease in productivity. Recently, workplace stressors have increased due to the COVID-19 pandemic. This study aims to contribute to the literature base in a couple of areas. First, it extends the current knowledge base by utilising generative additive modelling (GAMs) to uncover the nature of the relationship between workload (a key workplace stressor) and productivity based on real-world event logs. Additionally, it uses recursive partitioning modelling to shed light on the factors that drive the relationship between these variables. Secondly, it utilises a simulation-based approach to investigate the diffusion of workload-induced stress in the workplace. Simulation is a valuable tool for exploring the effect of changes in a risk-free manner as it provides the ability to run multiple scenarios in a safe and virtual environment with a view to making recommendations to stakeholders. However, there are several recognised issues with traditional simulation approaches, such as inadequate resource modelling and the limited use of simulations for operational decision making. In this study, we propose an approach which extracts the required parameters from an event log and subsequently utilises them to initialise a workload-induced stress diffusion simulation model accurately. We also explore the effects of varying the parameters to control the spread of workload-induced stress within the network. With suitable amendments, this approach can be extended to model the spread of disease (e.g., COVID-19), diffusion of ideas, among other things, in the workplace.

Keywords: process mining; simulation; workload-induced stress; Yerkes–Dodson Law



Citation: Ogunbiyi, N.; Basukoski, A.; Chausalet, T. Investigating the Diffusion of Workload-Induced Stress—A Simulation Approach. *Information* **2021**, *12*, 11.
<https://dx.doi.org/10.3390/info12010011>

Received: 2 December 2020
Accepted: 23 December 2020
Published: 26 December 2020

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Work-induced stress is defined as “the change in one’s physical or mental state in response to workplaces that pose an appraised challenge or threat to that employee” [1]. The impact of workplace stress includes “increased absenteeism, organizational dysfunction, and decreased work productivity” [1]. Workplace stress has also been linked to higher levels of alcohol consumption during retirement [2]. A key stressor in the workplace is the workload and pace of work [3]. Numerous studies have explored the relationship between workload and productivity. For example, Ref. [4] proposed a quadratic relationship between arousal (a proxy for workload-induced stress) and performance (see Figure 1).

Other studies have built on and extended this relationship, referred to as the Yerkes–Dodson law. For example, Ref. [5] describe the widely accepted explanation of the relationship. The authors posit that when the workload is below the optimal level of arousal and performance, performers are not as alert and hence do not perform at the optimal level. However, as the workload increases, so does alertness until the optimum level of performance is reached. Any increase in workload past this point results in decreased performance as performers “need more time to process information, to take decisions and, due to the high level of arousal, might make more mistakes”. They argue for a “load-based work order release” system which feeds work into the system based on the

existing workload on the shop floor and posit that this has a positive impact on increasing work order throughput times. Ref. [6] also explores this relationship utilising a process mining approach. Process mining describes a set of “techniques exploiting the information recorded in audit trails, transaction logs and databases” to “discover processes, check conformance, analyse bottlenecks and suggest improvements” [7,8]. Whilst [6] concluded that “the relationship described by the Yerkes Dodson law of arousal really exists”, the study stopped short of demonstrating the existence of the inverse U-shape relationship arguing that “more sophisticated . . . techniques” were required to confirm this.

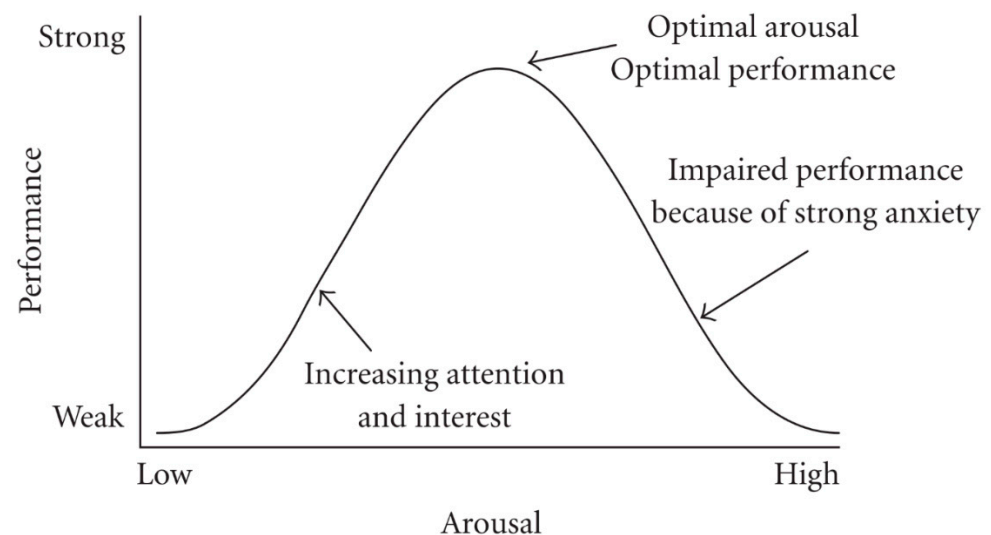


Figure 1. Hebbian version of the Yerkes–Dodson law (see [9]).

That gap is what the first half of this paper attempts to address. Utilising a couple of real-world event logs, we build generative additive models (GAMs) to uncover the nature of the relationship between workload and productivity. GAMs enable us to fit non-linear relationships to the data of interest and are relatively interpretable. In addition, we build a couple of recursive partitioning models to shed light on the factors that drive the relationship between these variables. Thus, this study contributes to the literature by uncovering the nature of the relationship between these variables and the factors that drive them.

The value of simulation as a way to rapidly explore the effect of changes in a risk-free manner has long been understood. However, Ref. [10] highlights several issues with traditional simulation approaches. Apart from the limited use of “existing artifacts such as historical data and workflow schemas”, the modelling of process performers is inadequate (e.g., the incorrect assumption that performers work at a constant speed or the assumption that performers immediately work on incoming tasks when they are available). To address these issues, Ref. [11] argues that “to adequately set these parameters and make sure that processes are modeled accurately, . . . the information available in event logs” needs to be exploited utilising process mining techniques. Ref. [10] highlights four pertinent types of data, namely: “event log” which describes historical information about recorded events, “process state” which represents information attached to cases, “process model” which describes the sequencing and routing of activities, and the “resource model” which conveys information about performers, roles, departments, etc.

Extending the link earlier established between workload and stress, the principle of emotional contagion—the phenomena of having one person’s emotions trigger emotions and related behaviours in others—has long been accepted [12,13]. More recent studies in the field of neuroscience have established the neurological basis of these phenomena [14,15]. We posit that as co-workers interact as they execute common activities simultaneously, stressed workers “infect” non-stressed workers and thus diffuse stress across the workplace.

To identify a suitable simulation model for the spread of workload-induced stress, we delve into the field of epidemiological research to examine models for exploring the transmission of infectious diseases. Ref. [16] propose a “general stochastic framework for modelling the spread of epidemics on networks”. This approach is an ideal choice for combining simulation with processing mining as there exist several studies which have successfully discovered social networks from event logs. For example, Ref. [17] proposes an approach for discovering social networks from an event log and several metrics based on potential causality, joint cases/activities, and special event types. They also apply these concepts to a real-life event log. In [18], the authors build on these and extend the approach to discover organisational models from event logs.

In the second half of this study, we discover a social (co-worker) network from an event log and utilise the network properties to initiate a simulation model which explores the spread of workload-induced stress. We further contribute to the literature base by proposing a novel approach which investigates the diffusion of workload-induced stress utilising an epidemiological simulation model initialised with parameters extracted from an event log. Whilst the focus in this study is the diffusion of workload-induced stress, with suitable amendments, the model can also be used to explore the spread of disease (e.g., COVID-19) in the workplace or the diffusion of ideas, amongst other things.

The remainder of the paper is structured as follows. Section 2 defines vital terms built on throughout the paper and describes the proposed approach, while Section 3 details the evaluation results of the proposed approach. The penultimate section summarises the findings and describes the threats to the validity of the study, while the final section suggests further research areas for extending these.

2. Materials and Methods

2.1. Definitions

2.1.1. Event, Traces, and Event Logs

Definition 1. “Event”. Let ε represents the event universe and T the time domain, A represents the set of activities, and P represents the set of performers (i.e., individuals and teams).

An event e is a tuple $(\#case_identifier(e), \#activity(e), \#start_time(e), \#completion_time(e), \#attribute_1(e), \dots, \#attribute_n(e))$. The elements of the tuple represent the attributes associated with the event. Though an event is minimally defined by the triplet $(\#case_identifier(e), \#activity(e), \#completion_time(e))$, it is common and desirable to have additional attributes such as $\#performer(e)$ indicating the performer associated with the event and $\#trans(e)$ indicating the transaction type associated with the event, amongst others. For each of these attributes, there is a function which assigns the attribute to the event. e.g., $attr_{start_time} \in \varepsilon \rightarrow T$ assigning a start time to the event, $attr_{completion_time} \in \varepsilon \rightarrow T$ assigning a completion time to the event, $attr_{activity} \in \varepsilon \rightarrow A$ assigning an activity label to the event, and $attr_{performer} \in \varepsilon \rightarrow P$, a partial function assigning a performer (or resource) to events. Note that $attr_{performer}$ is a partial function as some events may not be associated with any performers.

An event is often identified by the activity label $(\#activity(e))$ which describes the work performed on a process instance (or case) that transforms input(s) to output(s).

Definition 2. “Trace”. A trace is a (time-increasing) sequence of events, $\sigma \in \varepsilon^*$ such that each event appears only once, i.e., for $1 \leq i < j \leq |\sigma|$: $\sigma_i \neq \sigma_j$ and $\sigma_i \leq \sigma_j$.

Definition 3. “Event log”. Let C represent the set of all traces, both full and partial. An event log is a set of traces $L \subseteq C$ for a particular process such that each event appears at least once in the log, i.e., for any $\sigma_1, \sigma_2 \in L$: $\forall e_1 \in \sigma_1 \forall e_2 \in \sigma_2$ $e_1 \neq e_2$ or $\sigma_1 = \sigma_2$.

Definition 4. “Processing time”. Let e represent an event, $\#start_time(e)$, the start time associated with the event, and $\#completion_time(e)$, the completion time associated with the event. The

processing time for e , $\tau_{proc} = \#completion_time(e) - \#start_time(e)$. It indicates the time taken to complete processing the event.

Definition 5. “Workload”. Let A represent the set of valid activity labels, W represent a time window with start, W_{start} , end, W_{end} , and event log L . The workload function is defined as:

$$workload(\#activity(e), W_{start}, W_{end}) \rightarrow \mathbb{N}$$

where \mathbb{N} is the set of natural numbers $\{0, 1, 2, 3 \dots\}$. This denotes the number of instances of a specific activity present in time window W .

We further define $attr_{completed}(p, a) \rightarrow \mathbb{N} \leq workload$ for each performer/activity pair to indicate the number of events (with activity label a) the performer p completed in the given time window.

Definition 6. “Average processing speed”. Given a set of valid activity labels A , a time window W , and a performer p , the processing speed is defined by:

$$s = \frac{\sum \tau_{proc} : \#start_time(e_i) \wedge \#completion_time(e_i) \in W}{completed(p, a)} \quad (1)$$

This indicates the average processing speed for the performer/activity pair in the given time window.

2.1.2. Social Networks and Network Models

Definition 7. “Co-worker network”. Let P represent the set of performers, E represents a set of undirected edges, and $\phi : E \rightarrow \{x, y : x, y \in P\}$ represent an incidence function mapping edges to vertices defined as follows:

$$= \begin{cases} 1, & \text{if } \#activity_label(e_i) = \#activity_label(e_j) \wedge \\ & [\#start_time(e_i), \#completion_time(e_i)] \cap [\#start_time(e_j), \#completion_time(e_j)] > 0 \\ 0, & \text{Otherwise} \end{cases} \quad (2)$$

A co-worker graph is an undirected multigraph $G = (P, E, \phi)$. For our study, the incidence function maps an edge when two performers are co-workers as well as the duration of each interaction. Two performer x and y are considered co-workers if x completes e_i , y completes e_j , with both events having identical activity labels and the processing time interval for both events overlaps (see Figure 2).

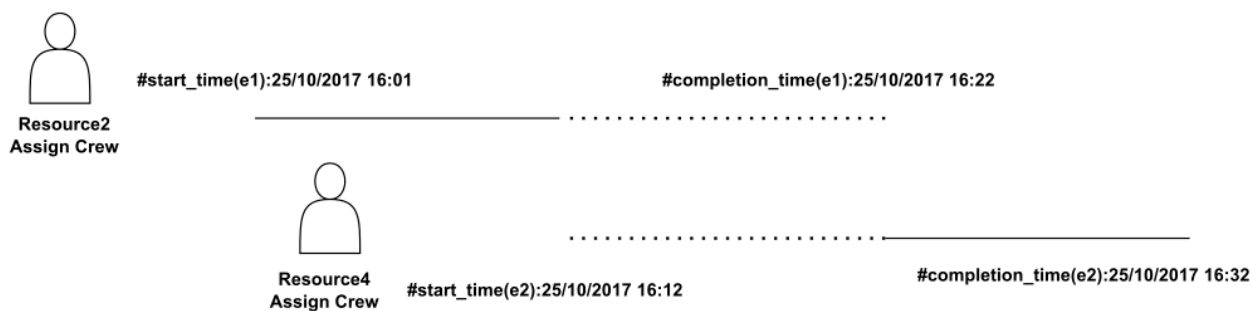


Figure 2. Co-worker network.

Definition 8. “Network model”. Let G denote an undirected graph representing the co-worker network, P represent the set of performers. The partnership formation process for the network simulation model is defined by:

$$\text{logit} [P(G_{xy,t+1} = 1 | G_{xy,t} = 0, G^c)] = \theta \frac{T}{+} \delta(g + (z)) \quad (3)$$

where G_{xy} denotes the edge between vertices $x, y \in P$, $P(G)$ denotes the probability distribution of the network, z denotes the observed network, G^c denotes the rest of the network, θ denotes the conditional log-odds of G_{xy} as a function of the number of configurations it creates and $\delta(g)$ denotes change statistics that indicate how the count of configurations change when G_{xy} is toggled from 0 to 1. Note that G_{xy} is indexed by time and formation at time $t+1$ is conditional on G_{xy} existing by time t [16].

The complimentary edge dissolution process is defined as follows:

$$\text{logit} [P(G_{xy,t+1} = 1 | G_{xy,t} = 0, G^c)] = \theta \frac{T}{-} \delta(g - (z)) \quad (4)$$

To illustrate the terms above, consider a process for reporting and remediating defects to public goods. A snippet of the event log is shown in Table 1. The set of valid activity labels is as follows: ‘Create Service Request’, ‘Initial Review’, ‘Assign Service Request’, ‘Assign Crew’, ‘Contact Citizen’, ‘Put Service Request On Hold’, ‘Close Service Request’.

Table 1. Event Log Example.

Service Request ID	Activity	Start Time	End Time	Performer
XY4567	Create Service Request	22/10/2017 18:34	22/10/2017 18:38	Citizen1
XY4567	Initial Review	25/10/2017 10:12	25/10/2017 10:14	Resource1
XY4567	Accept Ownership	25/10/2017 10:16	25/10/2017 10:17	Resource1
XY4567	Assign Service Request	25/10/2017 11:26	25/10/2017 11:29	Resource1
XY4567	Assign Crew	25/10/2017 16:01	25/10/2017 16:22	Resource2
XY4567	Contact Citizen	27/10/2017 11:04	27/10/2017 11:09	Resource2
XY4567	Close Service Request	27/10/2017 11:45	27/10/2017 11:55	Resource2
XY8910	Create Service Request	21/10/2017 15:12	22/10/2017 15:20	Citizen2
XY8910	Accept Ownership	22/10/2017 11:22	25/10/2017 11:25	Resource3
XY8910	Assign Crew	25/10/2017 16:12	25/10/2017 16:32	Resource4
XY8910	Close Service Request	26/10/2017 12:23	26/10/2017 12:55	Resource4

Figure 2 illustrates the concept of “co-workers” (see Definition 7). We observe that the processing time interval for #case_identifier(XY4567), #activity(Assign Crew) executed by Resource2 overlaps by 10 min with the processing time interval for #case_identifier(XY8910), #activity(Assign Crew) executed by Resource4 (see bold font). Thus an edge is formed between Resource2 and Resource4 in the co-worker network, and the duration of interaction (or exposure) is 10 min (see dashed lines).

2.2. Evaluation

In this section, we describe the two sets of analyses performed to address the research questions of interest in this study. In the first set of analyses, we evaluate the relationship

between the workload and processing speed to determine whether it displays a quadratic relationship (as predicted by the Yerkes–Dodson law) and if so, under which conditions. In the second set of analyses, we simulate a network model to investigate the diffusion of workload-induced stress in a co-worker network. Specifically, we seek to address the following research questions:

RQ1: Does the relationship between workload and processing speed exhibit a quadratic relationship as proposed by the Yerkes–Dodson law?

RQ2: If so, when does this relationship hold and when not?

RQ3: Do network simulation approaches facilitate the discovery of successful interventions to mitigate the diffusion of workload-induced stress?

In the following section, we provide further details about the setup and how we answer the research questions.

2.2.1. Datasets

Two real-life event logs from the Business Process Intelligence Challenge (BPIC) were used as follows: BPIC12(W) [19], BPIC17(W) [20]. BPIC 12 contains event log data for a credit (i.e., personal loan or overdraft) application process at a Dutch financial institution. BPIC 17 contains data from the same process and institution; however from a different supporting system. These logs were selected as they contained a significant proportion of cases with both event start (`#start_time(e)`) and completion (`#completion_time(e)`) timestamps. This enabled us to calculate the processing speed for these events

For the simulation exercise, we used a synthetic event log [21] which contains the details for a repair process. This log was selected as, not only did it contain data which enable calculation of processing speed, but also information about the role which was used in the initialisation of the network simulation model.

See Table 2 for a summary of the logs used for the experiments.

Table 2. Event Log Overview.

	BPIC 17(W)	BPIC 12(W)	Repairs Log
Number of events	768,823	170,107	15,486
Number of cases	31,509	9658	1104
Number of traces	10,701	2643	80
Number of distinct activities	8	7	8
Mean trace length	24.40	17.61	14.03
Mean throughput time (days)	21.89	11.68	0.05
Throughput time SD (days)	13.17	12.79	0.01
Domain	Financial services	Financial services	IT Support

2.2.2. Experimental Setup

To investigate the first two research questions, we implemented a function in R to calculate the daily workload and average processing speed for each performer (*time window start* = 00:00:00; *time window end* = 23:59:59). We selected this window for the sake of parsimony and due to the presence of activities in the log which complete late in the day (e.g., after 23:00). Hence we decided not filter the log to a typical workday (i.e., 08:00–18:00) as the observed work pattern did not fit this. We considered calculating the daily workload and processing speed per performer better to capture the true nature of the demand on performers. However, we realised that the mean for different activities differed based on activity complexity; as such, combining all the activities performers had completed each day was likely to distort the average processing speed. As such, we adopted the methodology used in [6] and calculated the total daily workload for each activity, the number of activities each performer completed daily, and the average processing speed per performer/activity. We also calculated the cumulative workload for each activity and the number of activities each performer completed over the event log. We subsequently fitted a generalised additive model (GAM) to uncover the relationship between total daily

workload and processing speed. We smoothed the GAMs with the restricted maximum likelihood method as this is widely acknowledged as most likely to produce stable and reliable results. We created a GAM model for each performer/activity combination with the average processing speed as the dependent variable and the total daily workload as the independent variable. In addition, we extracted relevant statistics from each model such as the expected degrees of freedom (edf) which indicates the complexity of the model's smooth, and the p -value, amongst others.

To answer the last research question, we created a social network from the co-worker network of performers in the Repairs event log. We subsequently extracted the following network properties from the co-worker network: the number of edges and vertices in the network, the number of stressed edges, the number of homogenous edges, the number of concurrent interactions, and the mean duration of interactions. Each performer was assigned an appropriate state (stressed/not stressed) based on whether their daily workload completed fell within the final quartile. As established earlier, based on findings in the literature, we posit that a stressed performer (i.e., infected) can spread stress through the co-worker network by “infecting” non-stressed performers via the process of emotional contagion. Borrowing from the field of epidemiology, we created a susceptible-infected-susceptible (SIS) model to simulate the diffusion of stress across the worker network. We chose this model (as opposed to a susceptible-infected-recovered (SIR) model which assumes immunity once recovered, for example) as a performer is again susceptible to workload-induced stress after recovery. The extracted co-worker network properties were used to initiate the SIS model. We concluded by executing multiple runs of the simulation to determine the effect that varying the infection probability and recovery rate had on the number of performers who were stressed at the end of the simulation run.

3. Results

3.1. Experimental Results

For the first set of results, to explore the relationship between workload and performance, we filtered for the GAM models which were significant at the 90% confidence level. Table 3 shows the distribution of the edf for the models. We observed that across both datasets, 43% of models had an edf of 1 indicating a linear relationship, 18% an edf of 2 indicating a quadratic relationship, and 40% an edf greater than or equal to 3, indicating a more complex smooth (see Figure 3a–d for example plots). We noted that there is partial support for the inverse U shape in the literature as even the more complex smooths (i.e., $\text{edf} \geq 3$) demonstrate this relationship. Note that the scale of the plots are shifted by the value of the intercept to aid interpretability. Hence we can predict the output assuming other variables are held at their average value. For example, for plot 3b, the predicted productivity for User 11009 performing activity “W_Completeren aanvraag” at the daily workload of 300 cases is 10 cases per day.

We explored further to uncover the factors which drive the nature of the relationship between the average processing speed and the daily workload. Borrowing from the approach adopted in [22], we built a couple of recursive partitioning model (rpart) from the GAM model data. The rounded edf for each model was selected as the classification target and the cumulative workload, activity label, and average processing speed were the independent variables. Figures 4 and 5 show the binary tree representation of the model. We expanded the GAM models to include all those significant at the 80% confident level to broaden the dataset.

Table 3. Distribution of effective degrees of freedom (edf) for generalised additive models (GAMs).

	1	2	≥ 3
BPIC 12 (W)	44%	17%	39%
BPIC 17 (W)	41%	18%	41%

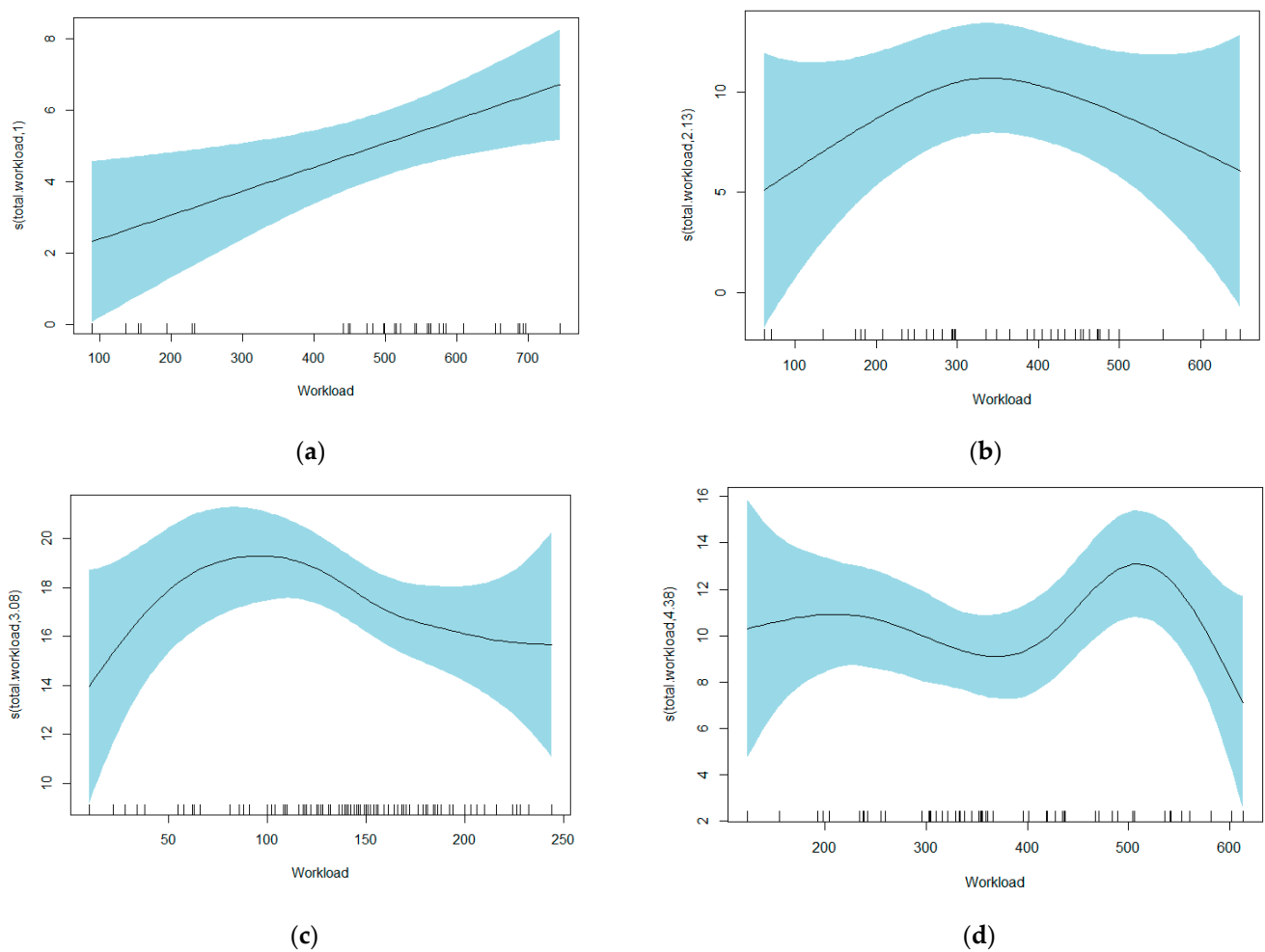


Figure 3. GAM plots of workload against processing speed. (a) BPIC 17 User 53/W_Validate application (edf = 1); (b) BPIC 12 User 11009/W_Completeren aanvraag (edf = 2); (c) BPIC 12 User 10972/W_Valideren aanvraag (edf = 3); (d) BPIC 17 User 24/W_Complete application (edf = 4).

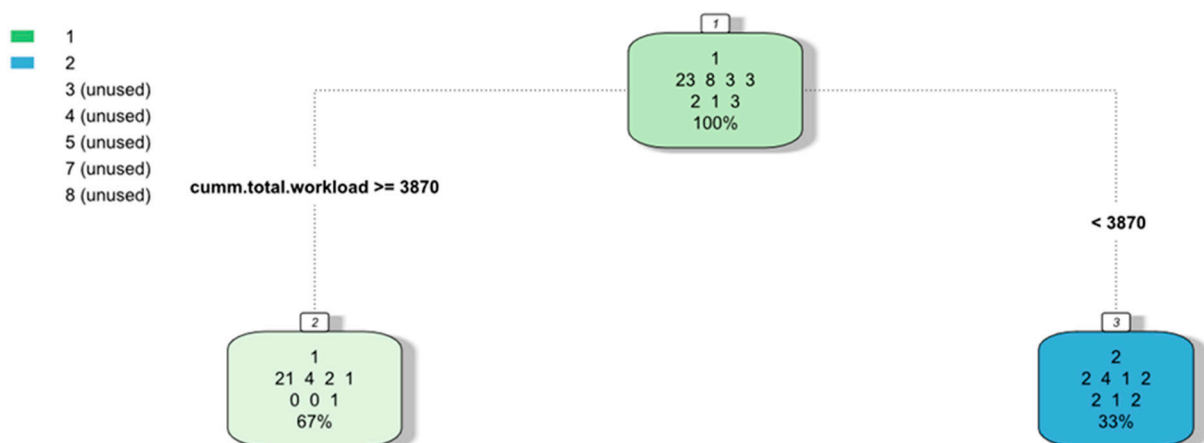


Figure 4. Business Process Intelligence Challenge (BPIC) 12—classification tree indicating factors driving degree of non-linearity (edf).

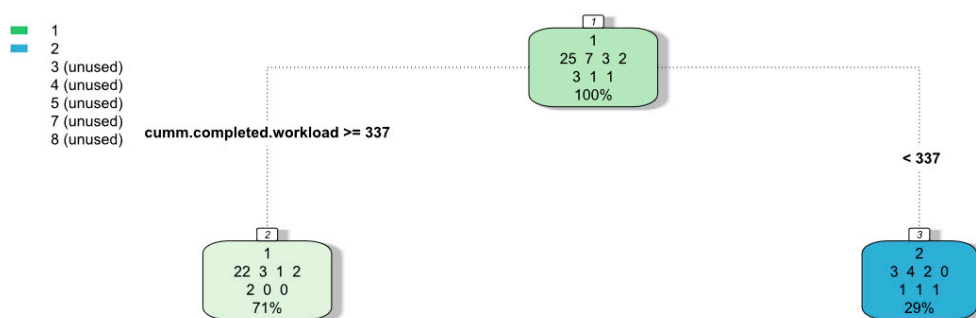


Figure 5. BPIC 17—classification tree indicating factors driving degree of non-linearity (edf).

We observed that for both datasets, the cumulative total workload or the cumulative number of cases completed by a performer were the factors that influenced whether the Yerkes–Dodson law is obeyed. Examining the rules that determine the classification for both datasets, it appears that the Yerkes–Dodson law is obeyed when a threshold value is surpassed; otherwise it is not. For the cumulative number of cases worked, the threshold value was the 28th percentile. Given that this attribute is a proxy for the individual experience of the performer, the results would seem to suggest that less experienced performers tend to obey the Yerkes–Dodson law. The other attribute is a proxy for the collective experience of the performers. Given that the threshold value is 32nd percentile, the results would seem to suggest that a less experienced workforce tends to obey the Yerkes–Dodson law.

These findings potentially have theoretical and practical implications. From a theoretical perspective, it potentially sheds light on the conditions under which the Yerkes–Dodson law applies and adds to the empirical basis on which the law is built. From a practitioner perspective, the findings have implications on work design, for example, in the design and implementation of an effective load-based work order release system. However, we recommend further research be undertaken to validate this with additional datasets and to test the generalisability of these results.

For the second set of results, we examined the effect of varying the infection probability and recovery rate on the number of performers who were stressed (i.num) at the end of the simulation run (time step 500). Figure 6 shows the plot of incidence, and recoveries for infection probability = 0.75 and recovery rate = 0.5. Table 4 shows the percentage of the workforce who was stressed at the end of the simulation run as the infection probability and recovery rate were varied.

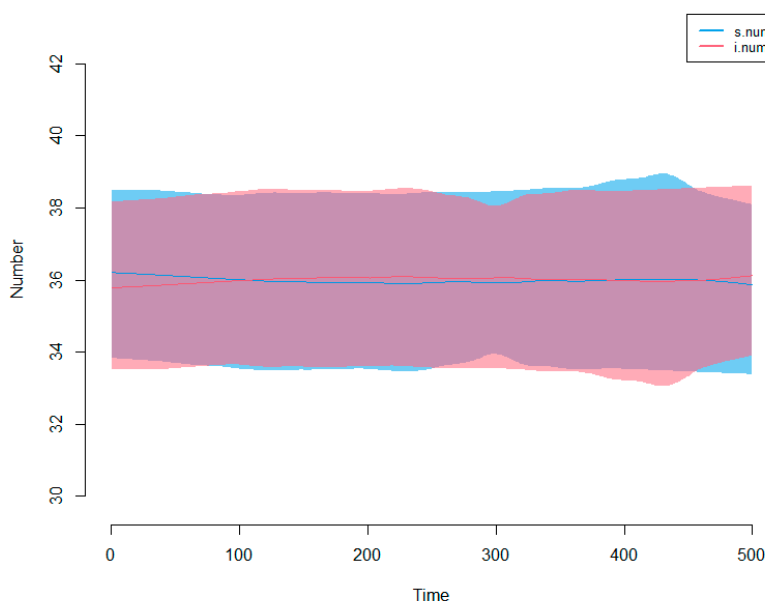


Figure 6. Stress simulation model—incidence and recoveries.

Table 4. Stress prevalence as a function of infection probability and recovery rate

Infection Probability	0.25	0.25	0.25	0.25	0.5	0.5	0.5	0.5	0.75	0.75	0.75	0.75	1	1	1	1
Recovery Rate	0.25	0.5	0.75	1	0.25	0.5	0.75	1	0.25	0.5	0.75	1	0.25	0.5	0.75	1
% Stressed (at t = 500)	76	50	25	0	74	50	25	0	78	49	28	0	75	50	25	0

We observed that the recovery rate had a more significant effect on reducing the prevalence of stress in the workplace, such that for all values of infection probability, at recovery rate = 1, the prevalence of stress in the workplace is eliminated. From a policy perspective, this finding potentially has implications for the allocation of stress management and intervention resources. Whilst intuitive reasoning might indicate that allocating more resources to interventions designed to reduce the infection probability are best (e.g., by making the workforce more resilient to “infection” by stress), the results would appear to indicate that interventions designed to increase the recovery rate (e.g., by engaging in moderate exercise, incorporating mindfulness techniques, etc.) are more effective [23,24] in reducing the prevalence of stress in the workplace.

It is worth noting that while there is a third model parameter that can be varied (act rate, which measures the mean number of interactions between co-workers), we chose not to vary this as we believe the adverse impact of reducing interaction (resulting in increased isolation and reduction in knowledge diffusion) outweighs any advantages gained by reducing stress prevalence

3.2. Threats to Validity

With regards to the first part of the study, we utilised two real-world datasets. These were the only real-world data that we were able to identify which contained the start and end timestamps for each activity, which was required to calculate the processing time. However, the relatively low number of datasets adversely impacts the ability to generalise these results. We would recommend repeating these experiments with additional real-world data sets to validate the results further

In addition, we highlight the propensity of GAMs and recursive partitioning models to overfit data as a limitation to these models as well as the added limitation that GAMs lose predictability when the independent variable is from a range outside of the observed data.

For our simulation model, we utilised a constant quantile applied against the total completed caseload for each performer to determine the stress status for the sake of parsimony. However, based on the results on the first part of the study, we realise that the determination of the stress status of each performer may differ (i.e., linear, quadratic, etc.) with different inflection points. We do not believe this significantly affected the results of the simulation as we visually examined the GAM plots for the performer/activity pairs to determine the optimal value of the appropriate quantile. However, we recognise that dynamically determining the optimal value for each performer/activity pair would be best.

Finally, our simulation model explored the effects of a single stressor (i.e., workload) in isolation. We recognise that in the real-world, multiple stressors exist in the workplace and they are likely to be in play simultaneously [3]. Our model does not consider these non-workload stressors and the interrelationships between them which is likely to impact the performance of the simulation model in a real-world setting (a known limitation of simulation models).

4. Conclusions

This study has attempted to uncover the nature of the relationship between workload (a key workplace stressor) and productivity from a couple of real-world event log utilising GAMs. We further explored the factors which drive this relationship. Whilst we found partial evidence for this law in the event log, this was in the minority, with the majority

of the relationships being linear. We also found that the cumulative total workload or the cumulative number of cases completed by a performer are factors that influence whether the Yerkes–Dodson law is obeyed, and that this happened when a threshold value was surpassed; otherwise it is not.

In the second part of the paper, we utilised a simulation-based approach to investigate the diffusion of workload-induced stress in the workplace. We found that in terms of stress management intervention, increasing the recovery rate yields better results vis-à-vis reducing the resilience of the workforce to stress.

In terms of further research, we propose several areas for further exploration. Firstly, we recommend that the study be repeated with different datasets and methodologies with a view to replicating the results and triangulating the conclusions. This would shed some more light on the generalisability of the results.

Secondly, in the event the thresholds we identified are replicated in further studies, we propose an exploration of why the identified factors (cumulative total workload or the cumulative number of cases completed by a performer) impact the Yerkes–Dodson law and why these thresholds occur where they do.

Thirdly, we propose the development of more sophisticated stress simulation models. For example, if the GAM model indicates a non-linear relationship between the workload and productivity, the simulation model could dynamically determine the inflection point at which productivity reduces. In addition, the model could simulate the impact of the prevalence of stress on variables of interest (e.g., on overall and individual productivity). Additionally, the model could factor in additional stressors such as the pace and variety of work and shift patterns—all of which can be derived from the event log—and model the interrelationships between them. Finally, we propose adding spatial context to the event log. This would enable us to calculate the distance between performers and better model the interaction between them. We can subsequently utilise these interactions to model the spread of information, disease, etc. in the workplace.

In future work, we intend to attempt to tackle a number of these opportunities.

Author Contributions: Conceptualisation, N.O.; methodology, N.O.; software, N.O.; validation, A.B. and T.C.; formal analysis, N.O.; investigation, N.O.; resources, N.O., A.B. and T.C.; data curation, N.O.; writing—original draft preparation, N.O.; writing—review and editing, A.B. and T.C.; visualisation, N.O.; supervision, A.B. and T.C.; project administration, N.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: No new data were created or analyzed in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Colligan, T.W.; Higgins, E.M. Workplace stress: Etiology and consequences. *J. Workplace Behav. Health* **2006**, *21*, 89–97. [\[CrossRef\]](#)
2. Richman, J.A.; Zlatoper, K.W.; Ehmke, J.L.Z.; Rospenda, K.M. Retirement and drinking outcomes: Lingered effects of workplace stress? *Addict. Behav.* **2006**, *31*, 767–776. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Bickford, M. *Stress in the Workplace: A General Overview of the Causes, the Effects, and the Solutions*; Canadian Mental Health Association Newfoundland and Labrador Division: St. John's, NL, Canada, 2005; p. 44.
4. Hebb, D.O. Drives and the CNS (conceptual nervous system). *Psychol. Rev.* **1955**, *62*, 243. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Bertrand, J.M.; Van Ooijen, H.P.G. Workload based order release and productivity: A missing link. *Prod. Plan. Control* **2002**, *13*, 665–678. [\[CrossRef\]](#)
6. Nakatumba, J.; van der Aalst, W.M. September. Analyzing Resource Behavior Using Process Mining. In *International Conference on Business Process Management*; Springer: Berlin/Heidelberg, Germany, 2009.
7. Van der Aalst, W.M.; Schonenberg, M.H.; Song, M. Time prediction based on process mining. *Inf. Syst.* **2009**, *36*, 450–475. [\[CrossRef\]](#)
8. Van der Aalst, W.M. *Process Mining: Data Science in Action*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2016.

9. Diamond, D.M.; Campbell, A.M.; Park, C.R.; Halonen, J.; Zoladz, P.R. The temporal dynamics model of emotional memory processing: A synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the Yerkes-Dodson law. *Neural Plast.* **2007**, *2007*, 60803. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Van der Aalst, W.M.; Nakatumba, J.; Rozinat, A.; Russell, N. *Business Process Simulation: How to Get It Right*; BPMcenter Org.: Brisbane, Australia, 2008; Volume 285, pp. 286–291.
11. van der Aalst, W.M. Business Process Simulation Revisited. In *Workshop on Enterprise and Organizational Modeling and Simulation*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 1–14.
12. Reik, T. *Listening with the Third Ear: The Inner Experience of a Psychoanalyst*; Farrar, Straus and Cudahy, Inc.: New York, NY, USA, 1948.
13. Jung, C.G. *Lecture Five: Analytical Psychology: Its Theory and Practice*; Random House: New York, NY, USA, 1968; pp. 151–160.
14. Iacoboni, M.; Molnar-Szakacs, I.; Gallese, V.; Buccino, G.; Mazziotta, J.C.; Rizzolatti, G. Grasping the intentions of others with one's own mirror neuron system. *PLoS Biol.* **2005**, *3*, e79. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Rizzolatti, G. The Mirror Neuron System and Imitation. *Perspect. Imitat. Neurosci. Soc. Sci.* **2005**, *1*, 55–76.
16. Jenness, S.M.; Goodreau, S.M.; Morris, M. EpiModel: An R Package for Mathematical Modeling of Infectious Disease over Networks. *J. Stat. Softw.* **2018**, *84*. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Van Der Aalst, W.M.; Reijers, H.A.; Song, M. Discovering Social Networks from Event Logs. *Comput. Supported Coop. Work* **2005**, *14*, 549–593. [\[CrossRef\]](#)
18. Song, M.; Van der Aalst, W.M. Towards comprehensive support for organisational mining. *Decis. Support Syst.* **2008**, *46*, 300–317. [\[CrossRef\]](#)
19. van Dongen, B.F. BPI Challenge 2012. 4TU. Centre for Research Data. Dataset. Available online: https://data.4tu.nl/articles/BPI_Challenge_2012/12689204 (accessed on 16 August 2020).
20. van Dongen, B.F. BPI Challenge 2017. Eindhoven University of Technology. Dataset. Available online: https://data.4tu.nl/articles/BPI_Challenge_2017/12696884 (accessed on 16 August 2020).
21. FutureLearn. Event Logs Used in This Course. Available online: <http://www.promtools.org/prom6/downloads/FutureLearn%20-%20Process%20mining%20with%20ProM%20-%20Event%20logs.zip> (accessed on 10 September 2020).
22. Hunsicker, M.E.; Kappel, C.V.; Selkoe, K.A.; Halpern, B.S.; Scarborough, C.; Mease, L.; Amrhein, A. Characterizing driver–response relationships in marine pelagic ecosystems for improved ocean management. *Ecol. Appl.* **2016**, *26*, 651–663. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Burton, J.P.; Hoobler, J.M.; Scheuer, M.L. Supervisor workplace stress and abusive supervision: The buffering effect of exercise. *J. Bus. Psychol.* **2012**, *27*, 271–279. [\[CrossRef\]](#)
24. Tetrick, L.E.; Winslow, C.J. Workplace stress management interventions and health promotion. *Annu. Rev. Organ. Psychol. Organ. Behav.* **2015**, *2*, 583–603. [\[CrossRef\]](#)